



UNIVERSITY OF LEEDS

This is a repository copy of *Structure and eigenvalues of heat-bath Markov chains*.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/89271/>

Version: Accepted Version

Article:

Dyer, M, Greenhill, C and Ullrich, M (2014) Structure and eigenvalues of heat-bath Markov chains. *Linear Algebra and Its Applications*, 454. 57 - 71. ISSN 0024-3795

<https://doi.org/10.1016/j.laa.2014.04.018>

© 2014. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Structure and eigenvalues of heat-bath Markov chains

Martin Dyer*

School of Computing
University of Leeds
Leeds LS2 9JT, UK
m.e.dyer@leeds.ac.uk

Catherine Greenhill†

School of Mathematics and Statistics
The University of New South Wales
Sydney NSW 2052, Australia
csg@unsw.edu.au

Mario Ullrich‡

Mathematisches Institut
Friedrich-Schiller-Universität
07743 Jena, Germany
mario.ullrich@uni-jena.de

10 April 2014

Keywords: stochastic matrices, Markov chains, heat-bath, eigenvalues, positive semidefinite
MSC 2010: 15B51, 60J10

Abstract

We prove that heat-bath chains (which we define in a general setting) have no negative eigenvalues. Two applications of this result are presented: one to single-site heat-bath chains for spin systems and one to a heat-bath Markov chain for sampling contingency tables. Some implications of our main result for the analysis of the mixing time of heat-bath Markov chains are discussed. We also prove an alternative characterisation of heat-bath chains, and consider possible generalisations.

1 Definitions and our first result

Suppose that Ω is a finite set and let $\pi : \Omega \rightarrow (0, 1]$ be a probability distribution on Ω . Let \mathcal{L} be a nonempty finite index set and let $L = |\mathcal{L}|$. Suppose that for all $x \in \Omega$ and $a \in \mathcal{L}$ we have a subset $\Omega_{x,a}$ of Ω such that

(I) $x \in \Omega_{x,a}$ for all $x \in \Omega$ and $a \in \mathcal{L}$, and

*Research supported by EPSRC Grant EP/J006300/1 and by a UNSW Faculty of Science Visiting Researcher Fellowship.

†Research supported by Australian Research Council.

‡Research partially supported by ERC Advanced Grant PTRELSS and DFG grant DFG GRK 1523.

(II) for each $a \in \mathcal{L}$, the set $\{\Omega_{x,a} : x \in \Omega\}$ forms a partition of Ω .

For $a \in \mathcal{L}$, define the $|\Omega| \times |\Omega|$ matrix P_a (with rows and columns indexed by Ω) by

$$P_a(x, y) = \frac{\pi(y)}{\pi(\Omega_{x,a})} \mathbb{1}(y \in \Omega_{x,a}). \quad (1)$$

(Here $\mathbb{1}(y \in \Omega_{x,a})$ equals 1 if $y \in \Omega_{x,a}$, and equals 0 otherwise.) Note that P_a is well-defined for all $a \in \mathcal{L}$, since π is nonzero on all states and each set $\Omega_{x,a}$ is nonempty.

Now for a given probability distribution ρ on \mathcal{L} , let P be the $|\Omega| \times |\Omega|$ matrix defined by

$$P = \sum_{a \in \mathcal{L}} \rho(a) P_a. \quad (2)$$

Since P is a stochastic matrix, it defines a Markov chain \mathcal{M} on Ω , determined uniquely by π , \mathcal{L} , ρ , and the sets $\Omega_{x,a}$. A transition of \mathcal{M} from current state $x \in \Omega$ is performed by choosing an element $a \in \mathcal{L}$ according to the distribution ρ , then sampling the next state y from $\Omega_{x,a}$ with respect to the distribution π restricted to $\Omega_{x,a}$.

Definition 1.1. A Markov chain \mathcal{M} on a finite state space Ω is a *heat-bath chain* if its transition matrix P satisfies (2) with respect to some finite nonempty set \mathcal{L} equipped with a probability distribution ρ , some probability distribution $\pi : \Omega \rightarrow (0, 1]$, and some sets $\Omega_{x,a}$ which satisfy (I), (II). Here the matrices P_a in (2) are defined by (1). \diamond

Note that conditions (I) and (II) imply that for all $x, y \in \Omega$ and $a \in \mathcal{L}$,

$$\text{if } y \in \Omega_{x,a} \text{ then } \Omega_{x,a} = \Omega_{y,a}. \quad (3)$$

Furthermore, when (2) holds it follows that \mathcal{M} is aperiodic (since every state has a self-loop) and that \mathcal{M} is reversible with respect to π . However, the chain \mathcal{M} need not be irreducible. (See [12] for Markov chain definitions which are not given here.)

Before proceeding, we indicate how our definition of heat-bath chains corresponds to the usual notion of heat-bath chains, in the setting of graph colourings or the Potts model. In such a chain, the state space is a subset of S^V for some finite sets V, S . To express this using our formulation, let $\mathcal{L} = \{a_1, \dots, a_L\}$ be the set of all those subsets $a_i \subset V$ which may be updated by a single transition of the chain, and, for $a \in \mathcal{L}$, let $\Omega_{x,a} = x_{V \setminus a} \times S^a$ be the set of all states which can be obtained from $x \in \Omega$ by “recolouring” or reassigning the values at elements of a . Here $x_W = (x_v)_{v \in W}$ denotes the restriction of x to W , for all $x \in S^V$ and $W \subseteq V$. So $\Omega_{x,a}$ contains all possibilities for the next state of the chain, given that x is the current state and that $a \in \mathcal{L}$ was chosen by the transition procedure. See also the examples presented in Section 2.

Lemma 1.2. Suppose that \mathcal{M} is a heat-bath chain, in the sense of Definition 1.1. Then \mathcal{M} has no negative eigenvalues.

Proof. By definition of P_a we know that $P_a(x, y) = 0$ if $y \notin \Omega_{x,a}$. Furthermore (3) implies that if $z \in \Omega_{x,a}$ then $P_a(x, y) = P_a(z, y)$ for all $y \in \Omega$. Therefore, for all $x, y \in \Omega$ and all $a \in \mathcal{L}$ we have

$$\begin{aligned} P_a^2(x, y) &= \sum_{z \in \Omega} P_a(x, z) P_a(z, y) = \sum_{z \in \Omega_{x,a}} P_a(x, z) P_a(z, y) = \sum_{z \in \Omega_{x,a}} P_a(x, z) P_a(x, y) \\ &= P_a(x, y). \end{aligned}$$

Hence $P_a^2 = P_a$, so P_a is an idempotent matrix. It follows that P_a is diagonalisable and the only eigenvalues of P_a are 0 and 1. (See for example [11, Section 3.3, Problem 3].)

Now let D be the diagonal $|\Omega| \times |\Omega|$ matrix with diagonal entries $(D)_{xx} = \sqrt{\pi(x)}$ for $x \in \Omega$. Define $Q_a = D^{-1}P_aD$ for all $a \in \mathcal{L}$. Since P_a is reversible with respect to π , it follows that Q_a is symmetric. Furthermore, Q_a is similar to P_a and hence has the same eigenvalues as P_a . Therefore Q_a is positive semidefinite, for all $a \in \mathcal{L}$. (Recall that a matrix is positive semidefinite if it is symmetric and has no negative eigenvalues.)

Now let $Q = \sum_{a \in \mathcal{L}} \rho(a) Q_a$. Since Q is a nonnegative linear combination of positive semidefinite matrices, it follows that Q is positive semidefinite. (See for example [11, Observation 7.1.3].) Furthermore, by definition we have $P = DQD^{-1}$, so P has the same eigenvalues as Q . Therefore P has no negative eigenvalues, as required. \square

1.1 Implications for the mixing time

Let \mathcal{M} be an ergodic, reversible Markov chain with finite state space Ω , transition matrix P and stationary distribution π . The eigenvalues of \mathcal{M} satisfy

$$1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N-1} > -1,$$

where $N = |\Omega|$. We refer to λ_{N-1} as the *smallest eigenvalue* of \mathcal{M} . The connection between the mixing time of a Markov chain and its eigenvalues is well-known (see [17, Proposition 1]):

$$\tau(\varepsilon) \leq (1 - \lambda_*)^{-1} \ln \frac{1}{\varepsilon \pi_{\min}} \quad (4)$$

where $\tau(\varepsilon)$ denotes the mixing time of the Markov chain, $\pi_{\min} = \min_{x \in \Omega} \pi(x)$ and

$$\lambda_* = \max\{\lambda_1, |\lambda_{N-1}|\}.$$

When studying the mixing time of a Markov chain \mathcal{M} using (4), the approach which has become standard is to make the chain \mathcal{M} *lazy* by replacing P by $(I + P)/2$, where I denotes the identity matrix. Then all eigenvalues of the lazy chain are nonnegative, and only the second-largest eigenvalue must be investigated. Clearly if P has no negative eigenvalues then $\lambda_* = \lambda_1$ and it is not necessary to make the chain lazy. Our result can be used to quickly verify this for heat-bath chains.

The bound (4) underpins many, but not all, methods of analysing the mixing time of a Markov chain. Heat-bath chains are often amenable to analysis using the classical technique of *coupling*, which is not based on (4). (As examples of coupling analyses of

heat-bath chains, see [1, 18].) For such chains, the information provided by Lemma 1.2 does not directly assist in bounding the mixing time.

However, in several applications including [3], a related heat-bath Markov chain is analysed using coupling, and then a comparison argument [4, 6] is applied to deduce rapid mixing of the original heat-bath chain. Comparison arguments typically relate the second-largest eigenvalues of the two chains, and hence they are often applied to lazy Markov chains. Lemma 1.2 demonstrates that it is unnecessary to make heat-bath chains lazy when applying the comparison method.

2 Two applications

In the special case that ρ is the uniform distribution over \mathcal{L} , the equation defining P is

$$P = \frac{1}{L} \sum_{a \in \mathcal{L}} P_a. \quad (5)$$

2.1 Application: a single-site heat-bath chain for spin systems

Let $G = (V, E)$ be an arbitrary graph and let S be a finite set of spins (or colors). Consider a state space $\Omega \subseteq S^V$ and let $\pi: \Omega \rightarrow (0, 1]$ be a probability distribution. Given $\sigma \in \Omega$, for all $v \in V$ and $k \in S$ we define $\sigma^{v,k}$ by

$$\sigma^{v,k}(u) = \begin{cases} \sigma(u) & \text{if } u \neq v, \\ k & \text{otherwise.} \end{cases}$$

(So $\sigma^{v,k}$ is obtained from σ by replacing the spin at v by k .) Additionally define, for $\sigma \in \Omega$ and $v \in V$, the set $S_v^\sigma = \{k \in S: \sigma^{v,k} \in \Omega\}$. (For spin systems with soft constraints, such as the Ising or Potts models, we have $\Omega = S^V$ and $S_v^\sigma = S$ for all $\sigma \in S^V$, $v \in V$.) The single-site *heat-bath chain* for Ω is the Markov chain with transition matrix defined by

$$P(\sigma, \tau) = \frac{1}{|V|} \sum_{v \in V} \frac{\pi(\tau)}{\sum_{\ell \in S_v^\sigma} \pi(\sigma^{v,\ell})} \mathbb{1}(\tau = \sigma^{v,\tau(v)})$$

for all $\sigma, \tau \in \Omega$. This matches the setting of Lemma 1.2 by choosing $\mathcal{L} = V$ and $L = |V|$, and defining

$$\Omega_{\sigma,v} = \{\sigma^{v,k} : k \in S_v^\sigma\}$$

for all $\sigma \in \Omega$ and $v \in V$. Hence, by Lemma 1.2, single-site heat-bath chains for general spin models do not have negative eigenvalues.

The heat-bath chain, which belongs to the family of Glauber dynamics (see for example [15]), has been studied by many authors including [10, 14, 16]. In several cases, the continuous-time version of this Markov chain is considered. One advantage of this approach is that mixing properties can be described solely by the second-largest

eigenvalue. Thus, when translating these results to discrete time, it usually remains to bound the smallest eigenvalue of the chain. The last example shows that for the heat-bath chain, the established continuous-time bounds can be used without further analysis. In the case of the Potts model this argument was used in the proof of [21, Theorem 2.10].

Note that there are other Glauber dynamics, such as the Metropolis chain, which are generally not guaranteed to have only nonnegative eigenvalues.

2.2 Application: a heat-bath chain for contingency tables

Let $\mathbf{r} = (r_1, \dots, r_m)$ and $\mathbf{c} = (c_1, \dots, c_n)$ be two vectors of positive integers with the same sum. A *contingency table* with row sums \mathbf{r} and column sums \mathbf{c} is an $m \times n$ matrix with nonnegative integer entries, such that the i 'th row sum is r_i and the j 'th column sum is c_j , for $i = 1, \dots, m$ and $j = 1, \dots, n$. Let $\Omega_{\mathbf{r}, \mathbf{c}}$ denote the set of all contingency tables with row sums \mathbf{r} and column sums \mathbf{c} .

Dyer and Greenhill [7] proposed a Markov chain for sampling contingency tables, which we will call the *contingency chain*. A transition of the chain is performed as follows: choose a 2×2 subsquare of the current table uniformly at random, then replace this 2×2 subsquare by a uniformly chosen 2×2 nonnegative integer matrix with the same row and column sums. The *lazy* contingency chain does nothing at each step with probability $\frac{1}{2}$, and otherwise performs a transition as described above. Cryan et al. [3] analysed the lazy contingency chain for a constant number of rows and proved that it is rapidly mixing.

To fit the contingency chain into the setting of (5), let \mathcal{L} be the set of all positions of 2×2 subsquares, and let $L = |\mathcal{L}| = \binom{m}{2} \binom{n}{2}$. Let P_a be the transition matrix of the Markov chain which acts only on the 2×2 subsquare $a \in \mathcal{L}$. Then $P_a(x, \cdot)$ is uniform over all contingency tables $y \in \Omega_{\mathbf{r}, \mathbf{c}}$ which differ from x only within the 2×2 subsquare a . Hence Lemma 1.2 applies (with π the uniform distribution on $\Omega_{\mathbf{r}, \mathbf{c}}$) and shows that the contingency chain has no negative eigenvalues.

3 A transfer result for positive semidefiniteness

The following result on matrices is well known. (The proof is easy, and omitted.)

Lemma 3.1. *Consider a state space Ω with probability distribution $\pi: \Omega \rightarrow (0, 1]$ and let P be a transition matrix on Ω . Let Ω' be a second state space with probability distribution $\mu: \Omega' \rightarrow (0, 1]$. Given any $|\Omega| \times |\Omega'|$ matrix R with rows indexed by Ω and columns indexed by Ω' , the adjoint R^* of R is defined by*

$$R^*(y, x) = \frac{\pi(x)}{\mu(y)} R(x, y) \quad \text{for all } x \in \Omega, y \in \Omega'.$$

Now suppose that $P = RTR^*$ where R and T satisfy the following conditions:

- R is a nonnegative $|\Omega| \times |\Omega'|$ matrix such that $\pi R = \mu$ and all rows of R sum to one, and
- T is a positive semidefinite transition matrix on Ω' which is reversible with respect to μ .

Then P is also positive semidefinite.

Note that we do not assume that T is irreducible. (If R is an invertible matrix and $P = RTR^*$ then P and T are often said to be *congruent*. But in our applications R need not be square.)

We now interpret the identity $P = RTR^*$ in terms of the corresponding Markov chains. Let \mathcal{M} be the Markov chain on Ω with transition matrix P , and let \mathcal{M}' be the Markov chain on Ω' with transition matrix T . A transition of \mathcal{M} from current state $x \in \Omega$ is performed as follows. First, generate an (auxiliary) state $x' \in \Omega'$ with respect to the probability distribution $R(x, \cdot)$. Then, perform one step of the chain \mathcal{M}' from initial state x' to obtain $y' \in \Omega'$. Finally, sample the new state $y \in \Omega$ of \mathcal{M} with respect to the distribution $R^*(y', \cdot)$.

Lemma 3.1 allows us to infer the positive semidefiniteness of P from the positive semidefiniteness of T . For some applications, Lemma 1.2 may be used to show that T is positive semidefinite, while in others we may argue more directly.

As an example, consider the *Swendsen-Wang chain* [20] for the q -state Potts model on a graph $G = (V, E)$. The state space is $\Omega = \{1, 2, \dots, q\}^V$ for some integer $q \geq 2$. For a fixed constant $\beta \geq 0$, the stationary distribution of the chain is defined by

$$\pi(\sigma) = Z^{-1} \exp\{\beta|E(\sigma)|\} \quad \text{for all } \sigma \in \Omega,$$

where

$$E(\sigma) = \{\{u, v\} \in E : \sigma(u) = \sigma(v)\}$$

denotes the set of monochromatic edges in σ , and Z is the normalizing constant. One step of this chain can be described as follows. Given the current state $\sigma \in \Omega$, sample a subset $A \subseteq E(\sigma)$ of the monochromatic edges such that each edge is included with probability $1 - e^{-\beta}$, with these choices all being independent. Then, colour each resulting connected component of the subgraph (V, A) with a new colour chosen from $\{1, \dots, q\}$ uniformly at random, with these choices all being independent. (For more details, see for example [8, 9, 21].)

This fits into the setting of Lemma 3.1 if we choose π, Ω as above, let

$$\Omega' = \{(\sigma, A) : \sigma \in \Omega, A \subseteq E(\sigma)\}$$

and define $\mu(\sigma, A) = Z^{-1}(e^\beta - 1)^{|A|}$ for all $(\sigma, A) \in \Omega'$. (This Z is the same normalising constant used to define π .) For all $\sigma \in \Omega$ and $(\tau, A) \in \Omega'$, let

$$R(\sigma, (\tau, A)) = e^{-\beta|E(\sigma)|} (e^\beta - 1)^{|A|} \mathbf{1}(\sigma = \tau).$$

Note that with this definition, $R^*((\sigma, A), \tau) = \mathbf{1}(\sigma = \tau)$ for all $\sigma \in \Omega$ and $(\tau, A) \in \Omega'$. Finally, for all $(\sigma, A), (\tau, B) \in \Omega'$, define

$$T((\sigma, A), (\tau, B)) = \mu((\tau, B) \mid B = A).$$

It is easy to verify that $\pi R = \mu$ and that $RT R^*$ equals the transition matrix of the Swendsen-Wang chain (see [8]).

Now observe that T is idempotent, and hence is positive semidefinite. We may also conclude this from Lemma 1.2, since T is a (rather trivial) heat-bath Markov chain in the sense of Definition 1.1 (where $\mathcal{L} = \{a\}$ has a unique element and setting $\Omega_{x,a} = \Omega$ for all $x \in \Omega$). Therefore Lemma 3.1 shows that the Swendsen-Wang chain has no negative eigenvalues, as claimed.

Another example of a Markov chain which fits the setting of Lemma 3.1 is the single-bond dynamics for the random-cluster model. Here Lemma 1.2 is needed in order to prove that the appropriate matrix T is positive semidefinite. See [21, Section 4.1] for more detail.

4 A characterisation of heat-bath chains

It follows from the proof of Lemma 1.2 that any nonnegative linear combination of stochastic idempotent matrices has only nonnegative eigenvalues. This leads us to ask whether Lemma 1.2 can be generalised to a wider class of Markov chains. To explore this question, we need some more definitions.

We use the symbols $\mathbf{0}, \mathbf{1}$ to denote any column vector or row vector with each entry equal to 0, 1 (respectively), of the appropriate size. We use symbols $\mathbf{a}, \mathbf{b}, \dots$ to denote column vectors, and use $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \dots$ to denote row vectors. Unless otherwise noted, the sizes of matrices and vectors can be inferred from the context.

A matrix M is called *substochastic* if it is nonnegative and $M\mathbf{1} \leq \mathbf{1}$. A square matrix M is called *permutation similar* to a matrix U if there is a permutation matrix A such that $U = A^T M A$. This operation corresponds to applying some permutation to both the rows and the columns of M to obtain U . Since $A^T = A^{-1}$ this is also a matrix similarity. We write $U \cong M$ to show that U and M are permutation equivalent (or $U \cong_A M$ to specify the permutation matrix A).

Note that M is reversible if and only if there is some nonnegative diagonal matrix D such that $MD = DM^T$. In particular, if all diagonal entries of D are positive then $D^{-1}MD = M^T$.

Remark 4.1. *The equivalence class of \cong_A is closed under multiplication and the taking of transposes, for all permutation matrices A . Hence if M is stochastic, idempotent or reversible, then so is any matrix U with $U \cong M$.*

We say that a matrix is an *SI matrix* if it is stochastic and idempotent. An *r-SI matrix* will refer to an SI matrix with rank r . We wish to obtain a characterisation of SI matrices. First we consider a generalisation of the stochastic case, which we will need later.

Lemma 4.2. *Let M be an irreducible, substochastic, idempotent matrix. Then M is a 1-SI matrix. Moreover, $M = \mathbf{1}\pi$, where π is a positive vector with $\pi\mathbf{1} = 1$.*

Proof. By idempotence, 0 and 1 are the only possible eigenvalues of M . Since M is irreducible, at least one eigenvalue of M is nonzero. This implies that M is stochastic, since otherwise M is irreducible and substochastic, but not stochastic: such matrices have spectral radius strictly less than one, see [11, Corollary 6.2.28]. Finally, using [11, Theorem 8.4.4], we obtain that 1 is a simple eigenvalue of M . This shows that M is a 1-SI matrix and hence is of the form $M = \mathbf{1}\pi$, where π is a positive vector and $\pi\mathbf{1} = 1$. \square

This immediately yields the following.

Corollary 4.3. *An SI matrix is irreducible if and only if it is a 1-SI matrix. Furthermore, a matrix M is 1-SI if and only if $M = \mathbf{1}\pi$ where π is positive and $\pi\mathbf{1} = 1$.*

A direct sum of 1-SI matrices is an SI matrix. However, an SI matrix need not be permutation equivalent to a direct sum of 1-SI matrices. Consider, for example, the matrix

$$M = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Clearly M is stochastic, and it is easy to check that $M^2 = M$, so M is idempotent. But M cannot be permuted to a direct sum of 1-SI matrices. This is due to the zero column in M , which corresponds to a state which is inaccessible from any state, including itself. Such a state $y \in \Omega$ is called *ephemeral* with respect to the Markov chain \mathcal{M} corresponding to M . Ephemeral states can only appear as the initial state of the chain.

The following characterisation of SI matrices depends on the number of ephemeral states.

Theorem 4.4. *Let M be a nonnegative square matrix. Then M is an SI matrix with exactly t zero columns if and only if $M \cong U$, where U has the form*

$$U = \begin{bmatrix} \mathbf{1}\pi_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{1}\pi_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \mathbf{1}\pi_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbf{1}\pi_k & 0 \\ \mathbf{p}_1\pi_1 & \mathbf{p}_2\pi_2 & \mathbf{p}_3\pi_3 & \cdots & \mathbf{p}_k\pi_k & 0 \end{bmatrix} \quad (6)$$

for some nonnegative vectors \mathbf{p}_i and positive vectors π_i which satisfy $\pi_i\mathbf{1} = 1$ for $i = 1, \dots, k$, and $\sum_{i=1}^k \mathbf{p}_i = \mathbf{1}$. (Here all diagonal blocks are square, though not necessarily of the same size: the last diagonal block has size $t \times t$.)

Proof. Suppose that M is an SI matrix with exactly t zero columns. Let M' be the matrix obtained from M by removing the t zero columns as well as the corresponding rows. Then M' is still stochastic, and $(M')^2 = M'$, so M' is an SI matrix with no zero columns. It is known [11, Section 8.3, Problem 8] that $M' \cong U'$, where

$$U' = \begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1k} \\ 0 & A_{22} & A_{23} & \cdots & A_{2k} \\ 0 & 0 & A_{33} & \cdots & A_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{kk} \end{bmatrix} \quad (7)$$

such that $A_{ij} \geq 0$ for $1 \leq i \leq j \leq k$, and A_{ii} is square and either irreducible or zero, for $i = 1, \dots, k$. Squaring U' gives

$$(U')^2 = \begin{bmatrix} A_{11}^2 & B_{12} & B_{13} & \cdots & B_{1k} \\ 0 & A_{22}^2 & B_{23} & \cdots & B_{2k} \\ 0 & 0 & A_{33}^2 & \cdots & B_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_{kk}^2 \end{bmatrix}, \quad (8)$$

for some $B_{ij} \geq 0$ ($i < j \leq k$). Hence we have $A_{ii}^2 = A_{ii}$ for $i = 1, \dots, k$. In particular $U_1 = A_{11}$ is idempotent, and U_1 is substochastic since U is stochastic. Since U has no zero column, $U_1 \neq 0$ and hence U_1 is irreducible. Therefore by Lemma 4.2 it follows that U_1 is stochastic, which implies that $A_{1j} = 0$ for $j = 2, \dots, k$. Thus $U' = U_1 \oplus U''$, where U_1 is a 1-SI matrix, and U'' is an SI matrix with no zero column, or is empty if $k = 1$ (in which case $U' = U_1$ is a 1-SI matrix). By induction, it follows that $U' = U_1 \oplus \cdots \oplus U_k$, where U_i is a 1-SI matrix for $i = 1, \dots, k$. Applying Corollary 4.3 shows that $U_i = \mathbf{1}\pi_i$, where π_i is a positive vector which sums to 1, for $i = 1, \dots, k$.

Hence we know that $M \cong U$ where

$$U = \begin{bmatrix} U_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & U_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & U_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & U_k & 0 \\ C_1 & C_2 & C_3 & \cdots & C_k & 0 \end{bmatrix}$$

for some nonnegative matrices C_1, \dots, C_k . Now $U^2 = U$, which implies that

$$C_i = C_i U_i = C_i \mathbf{1}\pi_i$$

for $i = 1, \dots, k$. Let $\mathbf{p}_i = C_i \mathbf{1}$, which is a nonnegative vector. Then $C_i = \mathbf{p}_i \pi_i$ and

$$\sum_{i=1}^k \mathbf{p}_i = \sum_{i=1}^k C_i \mathbf{1} = [C_1 \ C_2 \ \cdots \ C_k] \mathbf{1} = \mathbf{1},$$

as U is stochastic. This completes the proof of the “only if” statement.

For the converse, it suffices to assume that M satisfies (6), by Remark 4.1. Then it is not difficult to check that M is a SI matrix. \square

Corollary 4.5. *A matrix M is an r -SI matrix if and only if $M \cong U$, where U has the form (6) with $k = r$.*

Proof. By Remark 4.1 it suffices to consider U . If U has the structure given in (6) then U has k groups of rows of the form $\boldsymbol{\rho}_i = [\mathbf{0} \cdots \boldsymbol{\pi}_i \cdots \mathbf{0}]$ for $i = 1, \dots, k$. Hence U has rank at least k . The other rows are of the form $[\alpha_1 \boldsymbol{\pi}_1 \cdots \alpha_k \boldsymbol{\pi}_k] = \sum_{i=1}^k \alpha_i \boldsymbol{\rho}_i$, for some nonnegative constants $\alpha_1, \dots, \alpha_k$. Hence all rows of U are linearly dependent on the vectors $\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_k$. Thus U has rank exactly k , proving that $k = r$.

Conversely, if M is a r -SI matrix then M is an SI matrix. Theorem 4.4 states that $M \cong U$, where U is given by (6). The argument above then implies that $k = r$. \square

We are mostly interested in Markov chains with no ephemeral states (that is, with no zero columns in their transition matrix), where the following result will be useful.

Corollary 4.6. *Let M be a nonnegative square matrix. The following are equivalent.*

- (i) M is an SI matrix with no zero columns,
- (ii) $M \cong U$, where U is the direct sum of 1-SI matrices,
- (iii) M is an SI matrix which is reversible with respect to some positive distribution: that is, $D^{-1}MD = M^T$ for some diagonal matrix D with all diagonal entries positive.

Proof. That (i) and (ii) are equivalent follows from Theorem 4.4 by setting $t = 0$. Next, suppose that M is an SI matrix which satisfies $D^{-1}MD = M^T$ for some diagonal matrix D with all diagonal entries positive. This identity implies that if M has a zero column then M also has a zero row, contradicting the fact that M is stochastic. Hence (iii) implies (i).

Finally, we will prove that (ii) implies (iii). Note that it suffices to assume that M is the direct sum of 1-SI matrices, by Remark 4.1. Hence we have $M = U_1 \oplus \cdots \oplus U_k$ for some $k \geq 1$, where $U_i = \mathbf{1}\boldsymbol{\pi}_i$ for some positive vector $\boldsymbol{\pi}_i$ for $i = 1, \dots, k$. Define the positive vector $D_i = \text{diag}(\boldsymbol{\pi}_i)$ for $i = 1, \dots, k$, and let $D = D_1 \oplus \cdots \oplus D_k$. Then all diagonal entries of D are positive and

$$DM D^{-1} = \bigoplus_{i=1}^k D_i U_i D_i^{-1} = \bigoplus_{i=1}^k (D_i \mathbf{1})(\boldsymbol{\pi}_i D_i^{-1}) = \bigoplus_{i=1}^k \boldsymbol{\pi}_i^T \mathbf{1}^T = M^T,$$

proving that M is reversible with respect to D . Hence (ii) implies (iii), completing the proof. \square

We can now establish the following characterisation of heat-bath Markov chains.

Theorem 4.7. *Let \mathcal{M} be a Markov chain on the finite state space Ω , which is reversible with respect to the probability distribution $\pi : \Omega \rightarrow (0, 1]$. Then \mathcal{M} is a heat-bath chain (in the sense of Definition 1.1) if and only if the transition matrix P of \mathcal{M} is a nonnegative linear combination of nonnegative SI matrices with no zero columns.*

Proof. Suppose that \mathcal{M} is a heat-bath matrix. Then M satisfies (2) for some finite set \mathcal{L} and some probability distribution ρ on \mathcal{L} . Recalling (1) we see that each P_a is stochastic, and the proof of Lemma 1.2 shows that each P_a is idempotent. Finally, note that P_a has no zero columns since $x \in \Omega_{x,a}$ for all $x \in \Omega$, which implies that $P(x, x) > 0$ for all $x \in \Omega$. Hence P has the required form.

For the converse, suppose that $P = \sum_{a \in \mathcal{L}} \rho(a) P_a$ for some finite set \mathcal{L} , where each P_a is a nonnegative SI matrix with no zero column and $\rho(a) \geq 0$ for all $a \in \mathcal{L}$. Since P is also stochastic, it follows that ρ is a probability distribution on \mathcal{L} . Fix $a \in \mathcal{L}$. Corollary 4.6 shows that P_a is permutation-equivalent to a direct sum of 1-SI matrices, which we refer to as blocks. For each $x \in \Omega$, let $\Omega_{x,a}$ be the set of all states which correspond to a row in the block of P_a containing x . (This set is well-defined as it does not depend on the ordering of the blocks.) It follows that $x \in \Omega_{x,a}$ for all $x \in \Omega$ and $a \in \mathcal{L}$, and that the sets $\{\Omega_{x,a} : x \in \Omega\}$ form a partition of Ω , for each $a \in \mathcal{L}$. Hence conditions (I), (II) of Section 1 hold.

Now let $B_{x,a}$ be the block of P_a which corresponds to the set $\Omega_{x,a}$. Then $B_{x,a}$ is a 1-SI matrix, so it equals $\mathbf{1}\pi_{x,a}$ for some positive vector $\pi_{x,a}$ which sums to 1. However, $B_{x,a}$ is reversible with respect to the distribution π restricted to $\Omega_{x,a}$. It follows that for all $y, z \in \Omega_{x,a}$ we have

$$P_a(z, y) = P_a(x, y) = \frac{\pi(y)}{\pi(\Omega_{x,a})},$$

since $B_{x,a}$ has exactly one stationary distribution. Using the block structure of P_a , it follows that P_a satisfies (1) for all $a \in \mathcal{L}$. Therefore \mathcal{M} is a heat-bath chain in the sense of Definition 1.1, completing the proof. \square

4.1 Chains with finite convergence

We now investigate a possible generalisation of the notion of an SI matrix.

Suppose that M is an $n \times n$ stochastic matrix and, for every nonnegative vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ such that $\alpha \mathbf{1} = 1$, there exists a positive integer m_α such that

$$\lim_{t \rightarrow \infty} \alpha M^t = \alpha M^{m_\alpha}. \quad (9)$$

(We remark that this condition implies that M is aperiodic.) Such matrices have been considered before, and correspond to *chains with finite convergence* [2, 13, 19].

Write $\alpha M^{m_\alpha} = \pi_\alpha$. Then $\pi_\alpha M = \pi_\alpha$, since

$$\pi_\alpha M = \lim_{t \rightarrow \infty} \alpha M^{t+1} = \lim_{t \rightarrow \infty} \alpha M^t = \pi_\alpha.$$

Hence the Markov chain corresponding to M converges in a finite number of steps from any initial distribution. This generalises the blocks $B_{x,a}$ of the SI matrices P_a , which converge to their stationary distribution after one step.

Let \mathbf{e}_j be the j th unit (row) vector for $j = 1, \dots, n$. (Note, this breaks with our convention of using greek letters for rows and roman letters for columns.) For ease of notation, write m_j and π_j rather than $m_{\mathbf{e}_j}$, $\pi_{\mathbf{e}_j}$, for $j = 1, \dots, n$. Then

$$\pi_{\alpha} = \lim_{t \rightarrow \infty} \alpha M^t = \sum_{j=1}^n \alpha_j \lim_{t \rightarrow \infty} \mathbf{e}_j M^t = \sum_{j=1}^n \alpha_j \pi_j.$$

Let $m = \max_i m_i$, so that

$$\mathbf{e}_j M^m = \mathbf{e}_j M^{m_j} M^{m-m_j} = \pi_j M^{m-m_j} = \pi_j \quad \text{for } j = 1, \dots, n.$$

Hence

$$\alpha M^m = \sum_{j=1}^n \alpha_j \mathbf{e}_j M^m = \sum_{j=1}^n \alpha_j \pi_j = \pi_{\alpha},$$

so we may take $m_{\alpha} = m$ for all α . Then, for any nonnegative integer δ ,

$$\mathbf{e}_j (M^{m+\delta} - M^m) = \pi_j M^{\delta} - \pi_j = 0$$

for $j = 1, \dots, n$, which implies that $M^{m+\delta} = M^m$. Taking $\delta = 1$ shows that the eigenvalues λ of M satisfy $\lambda^m(\lambda - 1) = 0$. Hence the only eigenvalues of M are 0 and 1. Taking $\delta = m$ gives $M^{2m} = M^m$, so M^m is idempotent.

If M is also reversible then more is true, as we prove below. (We remark that the matrices C_j which appear in the statement of Lemma 4.8 are more general than those which arise in the proof of Theorem 4.7.)

Lemma 4.8. *Let M be a stochastic matrix with t zero columns. Suppose that there exists a positive integer m such that*

$$M^{m+\delta} = M^m \quad \text{for all } \delta \in \mathbb{N}.$$

Then $M \cong U$ for some matrix U with the block structure

$$U = \begin{bmatrix} U_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & U_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & U_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & U_k & 0 \\ C_1 & C_2 & C_3 & \cdots & C_k & 0 \end{bmatrix} \quad (10)$$

where the U_i are 1-SI matrices, the C_i are nonnegative, for $i = 1, \dots, k$, and the last diagonal block has size $t \times t$. In particular, if M is reversible with respect to some positive distribution then M is idempotent and, necessarily, $t = 0$.

Proof. By reordering the elements of Ω , we obtain a matrix U such that $M \cong U$, where U has the block structure

$$U = \begin{bmatrix} U_1 & A_{12} & A_{13} & \cdots & A_{1k} & 0 \\ 0 & U_2 & A_{23} & \cdots & A_{2k} & 0 \\ 0 & 0 & U_3 & \cdots & A_{3k} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & U_k & 0 \\ C_1 & C_2 & C_3 & \cdots & C_k & 0 \end{bmatrix}$$

such that A_{ij} is a nonnegative matrix for $1 \leq i < j \leq k$, while U_i is square, irreducible and substochastic and C_i is nonnegative, for $i = 1, \dots, k$.

Now each U_i is substochastic and irreducible, so U_i^m is substochastic, irreducible and idempotent. Therefore U_i^m is a 1-SI matrix, by Lemma 4.2. Hence U_i is stochastic: if the q 'th row sum of a substochastic matrix is strictly less than 1, then the same is true for any power of that matrix. It follows that $A_{ij} = 0$ for $1 \leq i < j \leq k$.

Furthermore, each U_i satisfies $U_i^m(U_i - I) = 0$, and hence has eigenvalue 1 (with multiplicity 1) with all other eigenvalues zero. It follows that U_i has rank 1, and since U_i is stochastic, this implies that $U_i = \mathbf{1}\pi_i$ for some positive vector π_i , for $i = 1, \dots, k$. By Corollary 4.3 it follows that U_i is a 1-SI matrix for $i = 1, \dots, k$, and (10) holds.

Finally, if M is reversible with respect to some positive distribution then M has no zero columns (that is, $t = 0$). Hence M is permutation equivalent to the direct sum of 1-SI matrices, and by Corollary 4.5 it follows that M is idempotent, as claimed. \square

Hence for reversible chains, there is no generalisation to Theorem 4.7 that can be obtained by replacing ‘‘idempotent’’ by some notion of finite convergence, as in (9).

Let m be the smallest integer such that $M^m = M$. Then M has complex eigenvalues if $m \geq 4$, which implies that any matrix satisfying this condition is not reversible. The case $m = 3$ corresponds to periodic Markov chains, which certainly have negative eigenvalues, and $m = 2$ is precisely the idempotence condition.

References

- [1] D. Achlioptas, M. Molloy, C. Moore and F. Van Bussell, Rapid mixing for lattice colourings with fewer colours, *Journal of Statistical Mechanics: Theory and Experiment* (2005), P10012.
- [2] I. Brosh and Y. Gerchak, Markov chains with finite convergence time, *Stochastic Processes and their Applications* **7** (1978), 247–253.
- [3] M. Cryan, M. Dyer, L.A. Goldberg, M. Jerrum and R. Martin, Rapidly mixing Markov chains for sampling contingency tables with a constant number of rows, *SIAM Journal on Computing* **36** (2006), 247–278.

- [4] P. Diaconis and L. Saloff-Coste, Comparison theorems for reversible Markov chains, *Annals of Applied Probability* **3** (1993), 696–730.
- [5] P. Diaconis and D. Stroock, Geometric bounds for eigenvalues of Markov chains, *Annals of Applied Probability* **1** (1991), 36–61.
- [6] M. Dyer, L.A. Goldberg, M. Jerrum and R. Martin, Markov chain comparison, *Probability Surveys* **3** (2006), 89–111.
- [7] M.E. Dyer and C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables, *Theoretical Computer Science* **246** (2000), 265–278.
- [8] S.F. Edwards and A.D. Sokal, Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm, *Phys. Rev. D* **38** (1988), 2009–2012.
- [9] V.R. Gore and M.R. Jerrum, The Swendsen Wang process does not always mix rapidly, *Journal of Statistical Physics* **97** (1999), 67–86.
- [10] T. Hayes, A simple condition implying rapid mixing of single-site dynamics on spin systems, in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, IEEE, Los Alimitos, 2006, pp. 39–46.
- [11] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [12] M. Jerrum, *Counting, Sampling and Integrating: algorithms and complexity*, Lectures in Mathematics – ETH Zürich, Birkhäuser, Basel, 2003.
- [13] B. Lindqvist, Ergodic Markov chains with finite convergence time, *Stochastic Processes and their Applications* **11** (1981), 91–99.
- [14] E. Lubetzky and A. Sly. Critical Ising on the square lattice mixes in polynomial time. *Comm. Math. Phys.*, 2012. doi:10.1007/s00220-012-1460-9.
- [15] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.
- [16] F. Martinelli and E. Olivieri, Approach to equilibrium of Glauber dynamics in the one phase region I. The attractive case, *Communications in Mathematical Physics* **161** (1994), pp. 447–486.
- [17] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow, *Combinatorics, Probability and Computing* **1** (1992), 351–370.
- [18] D. Štefanovič and E. Vigoda, Fast convergence of Markov chain Monte Carlo algorithms for phylogenetic reconstruction with homogeneous data on closely related species, *SIAM Journal on Discrete Mathematics* **25** (2011), 1194–1211.

- [19] E.J. Subelman, On the class of Markov chains with finite convergence time, *Stochastic Processes and their Applications* **4** (1976), 253–259.
- [20] Swendsen, R. H. and Wang, J.-S., Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.* **58** (1987), 86–88.
- [21] M. Ullrich, Rapid mixing of Swendsen-Wang dynamics in two dimensions, *to appear in Dissertationes Math.* (2014).